Table 3. Spatial Blurring Features For VTC/VT Imagery Of Figure 9

Scene	M-SI	SD-SI	RMS-SI	NPGT-SI
Top Row (NTSC)	70.4	103.3	125.1	14388
Second Row (DS1)	61.7	90.0	109.1	10247
Third Row (1/2 DS1)	60.9	89.5	108.3	9972
Bottom Row (1/4 DS1)	59.2	83.8	102.6	8265

2.6 Blocking, Edge Busyness, and Image Persistence Features

Blocking, defined in Table 1, is a severe form of spatial resolution degradation that normally occurs at low codec bit rates when there is a lot of motion in some sub-region or all of the video scene (such as during camera pans or zooms). Edge busyness and image persistence, also defined in Table 1, are video coding artifacts that causes false activity to appear around edges or elsewhere is the video scene. Blocking, edge busyness, and image persistence are most noticeable when the motion involves a high contrast (sharp) edge. Blocking, edge busyness, and image persistence cause edge energy to appear in the output video scene that was not present in the original input video scene. Human viewers have semantic knowledge of how certain items should look and they take objection to the presence of erroneous, out of place artifacts such as blocking, edge busyness, and image persistence. In particular, the appearance of false regular edge energy such as blocking is very noticeable and objectionable to the human viewer (more so than spatial blurring). Therefore, it is desirable to have a set of features that only measures the amount of false edge energy in a video scene.

Section 2.6.1 proposes a technique for extracting a set of features that quantitatively measures the amount of false edge energy in the output video scene. The features may be used to measure blocking, edge busyness, and image persistence since all contribute false edge energy to the output video scene. A by-product of the false edge energy feature is another set of features for measuring spatial blurring. The new

measures for spatial blurring, to be described below, do not contain false edge energy. Thus, the new measures for spatial blurring are more accurate than the measures presented in section 2.5.1 (M-SI, SD-SI, RMS-SI, NPGT-SI) when there is a large amount of false edge energy present in the video scene. The ability to separately measure spatial blurring and false edge energy may be important since each has a unique affect on perceived video quality. A video quality assessment system that detects the presence of blocking, a very objectionable artifact, could heavily penalize the overall quality rating.

2.6.1 Feature Extraction Technique

Edges in output video that are less intense than the corresponding edges in the input video are considered to have been blurred. False edge energy, due to the presence of such artifacts as blocking, edge busyness, or image persistence appears in the distorted output video but not in the undistorted input video. Suppose one were to compute an edge error image by subtracting the edge filtered output image from the corresponding edge filtered input image. Then, positive pixel values would be obtained for blurred output edges since, by definition, the edges in the input image are more intense (higher value) than the corresponding edges in the output image. Likewise, negative pixel values would be obtained for false output edges, since false edges in the output image are more intense than the corresponding edges in the input image. Thus, one could form an edge error image in which positive error represents blurring and negative error represents false edges. The exact feature extraction technique is given below.

1. Video alignment

   Multi-frame temporal alignment of the input and output video is required since the edges in the input and output images must be properly aligned. Thus, the extracted features will be representative of the "snapshot" performance of the video system under test since one is always comparing the output image with the closest matching input image. The single-frame temporal alignment method is not used because feature errors due to alignment could be generated, particularly if fields and/or frames have been omitted in the output video.

2. Video preconditioning

The sampled video imagery is preconditioned as previously described in section 2.5.1.

3. Edge extraction

An edge extraction filter is applied to the preconditioned video imagery as previously described in section 2.5.1. Here, a Sobel edge extraction filter was used (described in Appendix B).

4. Difference image

For each output/input video frame pair of interest, the Sobel difference image is computed as the Sobel filtered input image minus the Sobel filtered output image.

5. Feature computation

Several features can be extracted from the Sobel difference image. Eight are suggested here. Four of the eight are blurring features since they are extracted from the positive pixel values of the Sobel difference image. The other four are false edge features since they are extracted from negative pixel values of the Sobel difference image. All eight features possess the same desirable properties of features as the spatial blurring features for single-frame temporal alignment.

a. The mean of the positive Sobel difference image (M-PSDI)

M-PSDI is computed as the summation of the positive image pixel values divided by the total number of pixels in the sub-regional area of the image. The total number of pixels in the sub-regional area is used as the divisor, rather than just the number of positive pixels, so that the total amount of blurring energy can be directly compared to the total amount of false edge

energy. See Appendix A, equation 7 for a mathematical definition of M-PSDI.

b. The standard deviation of the positive Sobel difference image (SD-PSDI)

SD-PSDI is computed as the square root of (the summation of the squares of the positive image pixel values divided by the total number of pixels, minus the square of M-PSDI). See Appendix A, equation 8 for a mathematical definition of SD-PSDI.

c. The root mean square of the positive Sobel difference image (RMS-PSDI)

RMS-PSDI is computed as the square root of (the summation of the squares of the positive image pixel values divided by the total number of pixels). See Appendix A, equation 9 for a mathematical definition of RMS-PSDI.

d. The number of pixels greater than a threshold of the positive Sobel difference image (NPGT-PSDI)

NPGT-PSDI is computed as the total number of pixels within any sub-regional area that exceed a fixed threshold. Advantages of this feature include the ability to measure the number of severely blurred pixels, and ease of computation. See Appendix A, equation 10 for a mathematical definition of NPGT-PSDI.

e. The mean of the negative Sobel difference image (M-NSDI)

M-NSDI is computed as the summation of the negative image pixel values divided by the total number of pixels in the sub-regional area of the image. See Appendix A, equation 11 for a mathematical definition of M-NSDI.

f. The standard deviation of the negative Sobel difference image (SD-NSDI)

SD-NSDI is computed as the square root of (the sum of the squares of the negative image pixel values divided by the total number of pixels, minus the square of M-NSDI). See Appendix A, equation 12 for a mathematical definition of SD-NSDI.

g. The root mean square of the negative Sobel difference image (RMS-NSDI)

RMS-NSDI is computed as the square root of (the sum of the squares of the negative image pixel values divided by the total number of pixels). See Appendix A, equation 13 for a mathematical definition of RMS-NSDI.

h. The number of pixels less than a threshold of the negative Sobel difference image (NPLT-NSDI)

NPLT-NSDI is computed as the total number of pixels within any sub-regional area that are less than a fixed threshold. Advantages of this feature include the ability to measure the number of pixels corrupted with severe false edges, and ease of computation. See Appendix A, equation 14 for a mathematical definition of NPLT-NSDI.

Normalization of the above eight features can be performed by dividing by the appropriate spatial blurring features of the undistorted input video (M-SI, SD-SI, RMS-SI, and NPGT-SI from section 2.5.1). Then the amount of blurring or false edges in the output video with respect to the input video is obtained. The thresholds for NPGT-PSDI and NPLT-NSDI determine the severity of the blurring or false edges that the user is interested in measuring and these thresholds do not have to be identical to each other nor to NPGT-SI. The choice of the three thresholds will determine the range of the normalized features.

2.6.2 Sample VTC/VT Results

For illustrative purposes, the four spatial blurring features and four false edge features (blocking, edge busyness, and image persistence) were extracted from sampled VTC/VT imagery of a moving black ring against a white background (motion was from left to right). The high contrast moving edges of the black ring provided a sufficiently complicated object for the VTC/VT codec under test to exhibit the blocking, edge busyness, and image persistence artifacts. The top right image in Figure 11 shows the median filtered output of the VTC/VT codec that was operating at a bit rate of 1/4 DS1. The top left image in Figure 11 is the corresponding original NTSC input image after median filtering (found by using multi-frame temporal alignment). Note that the codec output image exhibits blurring, blocking, edge busyness, and image persistence (see Table 1). The bottom left and right images of Figure 11 show the Sobel edge extracted imagery of the NTSC input and codec output, respectively.

Figure 12 shows the Sobel difference image found by subtracting the bottom right image of Figure 11 from the bottom left image of Figure 11. For display purposes only (not for feature value computations), Sobel difference image pixel values were linearly scaled such that pixel values of zero (no error) are shown as a gray shade of 128 in Figure 12. The gray shade of 128 fell halfway between black (0) and white (255) on the 8 bit video printer that was used to generate the image. Thus, pixels that appear white are due to blurred edges in the output and pixels that appear black are due to false edges in the output. Clearly, the blurring energy has been separated from the blocking, edge busyness, and image persistence energy.

The four unnormalized blurring features and four unnormalized false edge features for Figure 12 are shown in Tables 4 and 5, respectively. To eliminate the erroneous edge energy at the image boundaries (due to median and Sobel filtering), all feature values were computed over a sub-rectangular region (size 504 horizontal pixels by 464 vertical pixels) centered on the main image. For NPGT-PSDI, pixel values that exceeded a threshold of 125 were counted. For NPLT-NSDI, pixel values that were less than a negative threshold of -125 were counted. For comparison to the reference, the spatial blurring features for the original NTSC Sobel extracted image (Figure 11, bottom left) were M-SI=14.7, SD-SI=27.7, RMS-SI=31.3, NPGT-SI=558 (calculated using a threshold of 250). Note that there was more blurring energy than false edge energy for this example.

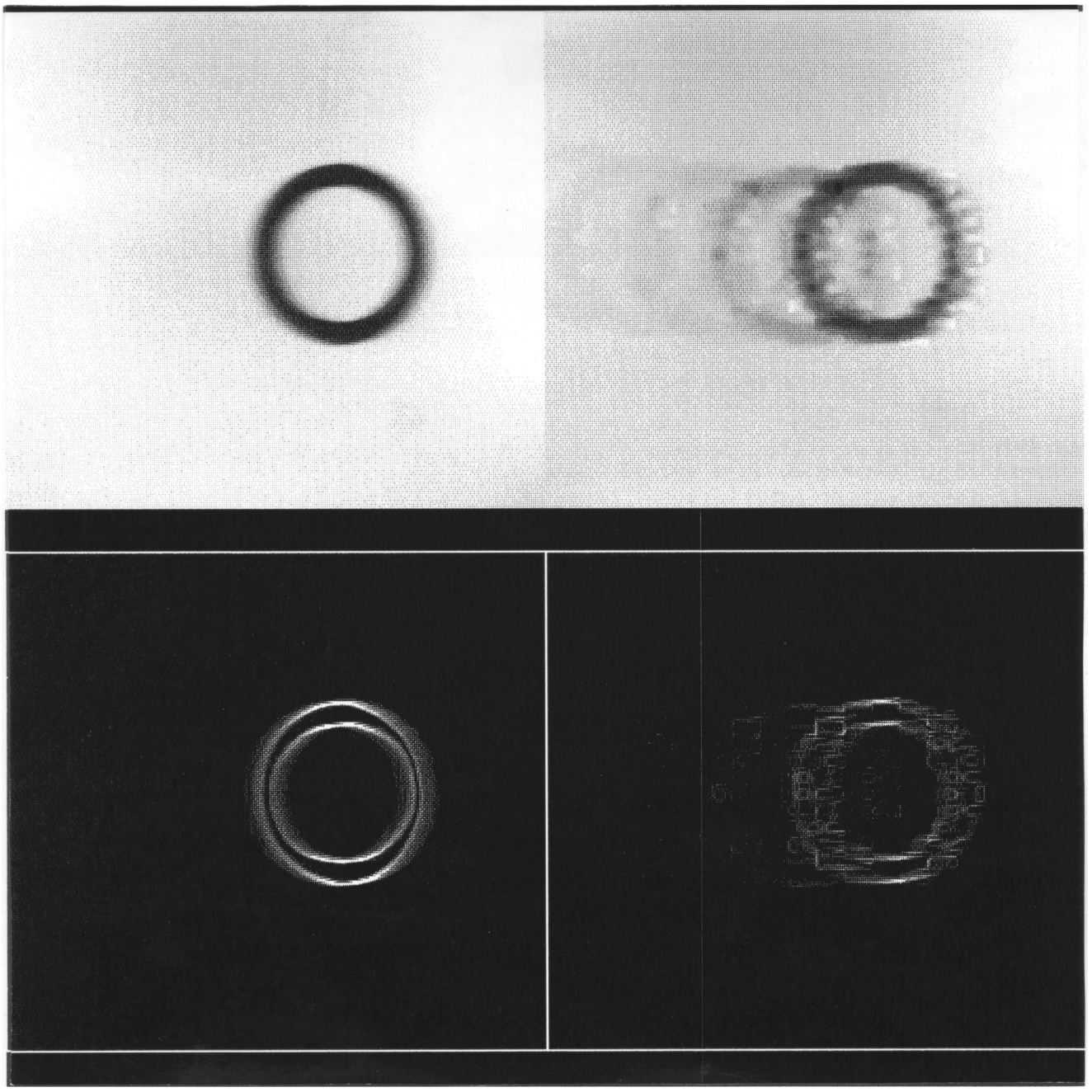

Figure 11. VTC/VT imagery of moving black ring against white background. Top left - median filtered NTSC input. Top right - median filtered codec output at rate 1/4 DS1. Bottom left - Sobel filtered NTSC. Bottom right - Sobel filtered codec output.

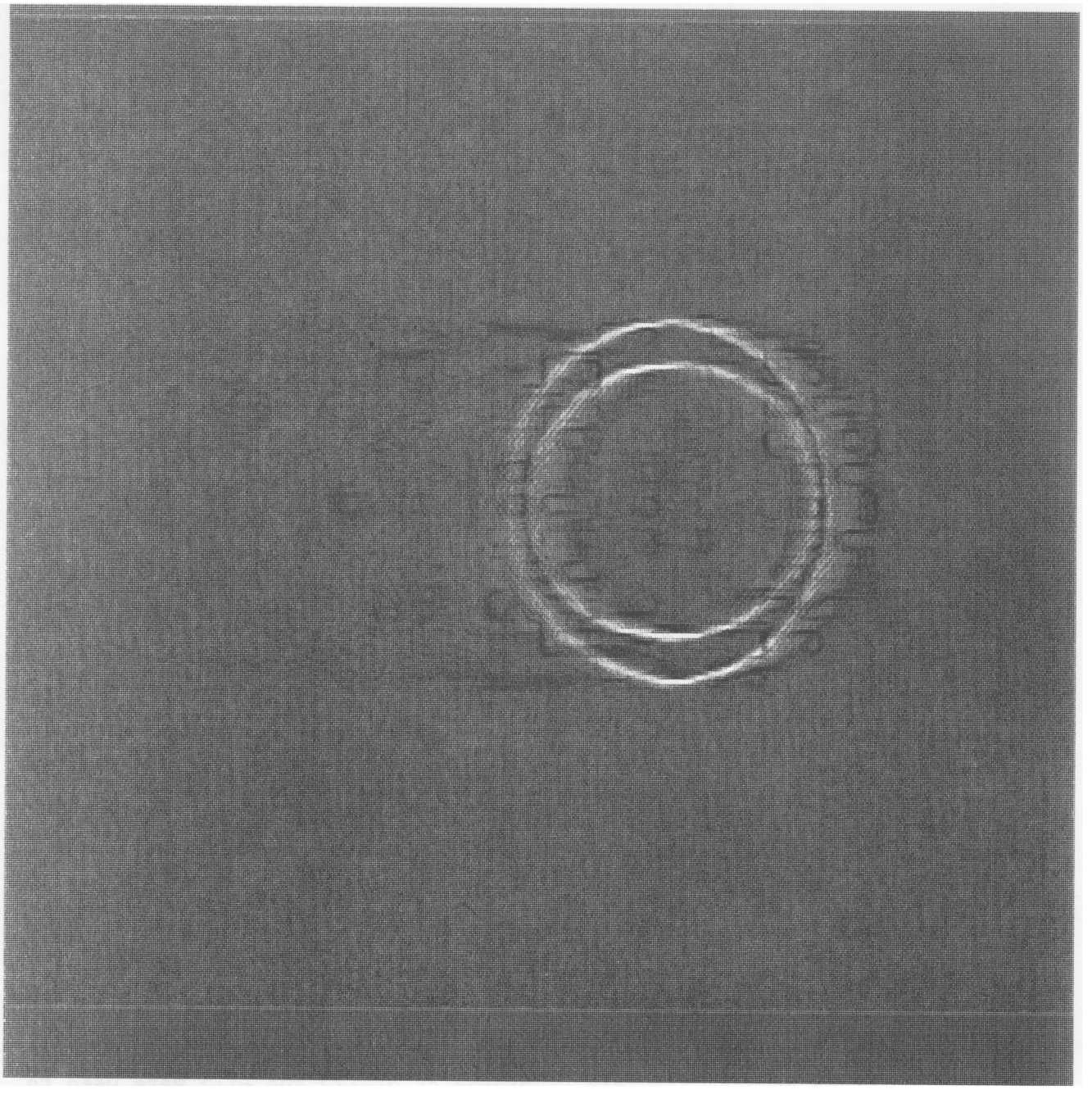

Figure 12. Sobel difference image of Figure 11. Note the separation of the blurring energy (white) from the blocking and edge busyness energy (black).

Table 4. Spatial Blurring Features For Figure 12

Scene	M-PSDI	SD-PSDI	RMS-PSDI	NPGT-PSDI
M o v i n g Ring (1/4 DS1)	5.6	15.9	16.8	882

Table 5. False Edge Features For Figure 12

Scene	M-NSDI	SD-NSDI	RMS-NSDI	NPLT-NSDI
M o v i n g Ring (1/4 DS1)	-5.1	9.3	10.6	26

Individual adjustment of the NPGT-PSDI and NPLT-NSDI thresholds can scale the importance of the blurring effects in relationship to the false edge effects.

Eight consecutive images of VTC/VT codec output video that contained upper body motion (the first four of which are shown in Figure 7) were processed to extract the spatial blurring, blocking, and edge busyness features. The corresponding NTSC input video frames for each of the codec output frames shown in rows 2, 3, and 4 of Figure 7 were found using multi-frame temporal alignment, rather than single-frame temporal alignment as shown in row 1 of Figure 7. Figure 13 shows the resulting Sobel difference images. The top, second, and bottom rows of Figure 13 were obtained by processing the codec output video for bit rates of DS1, 1/2 DS1, and 1/4 DS1, respectively. Tables 6 and 7 give the average of the unnormalized features, where the average was computed over 8 consecutive images at each bit rate. Thresholds of 125 and -125 were used to compute NPGT-PSDI in Table 6 and NPLT-NSDI in Table 7. The features in Tables 6 and 7 are directly comparable to the features in Table 3. Note from Figure 13 and Tables 6 and 7 that both blurring (white) and false edges (black) increased as the coding bit rate fell from DS1 to 1/2 DS1 to 1/4 DS1. Also note that there was more blurring than false edges at all bit rates.

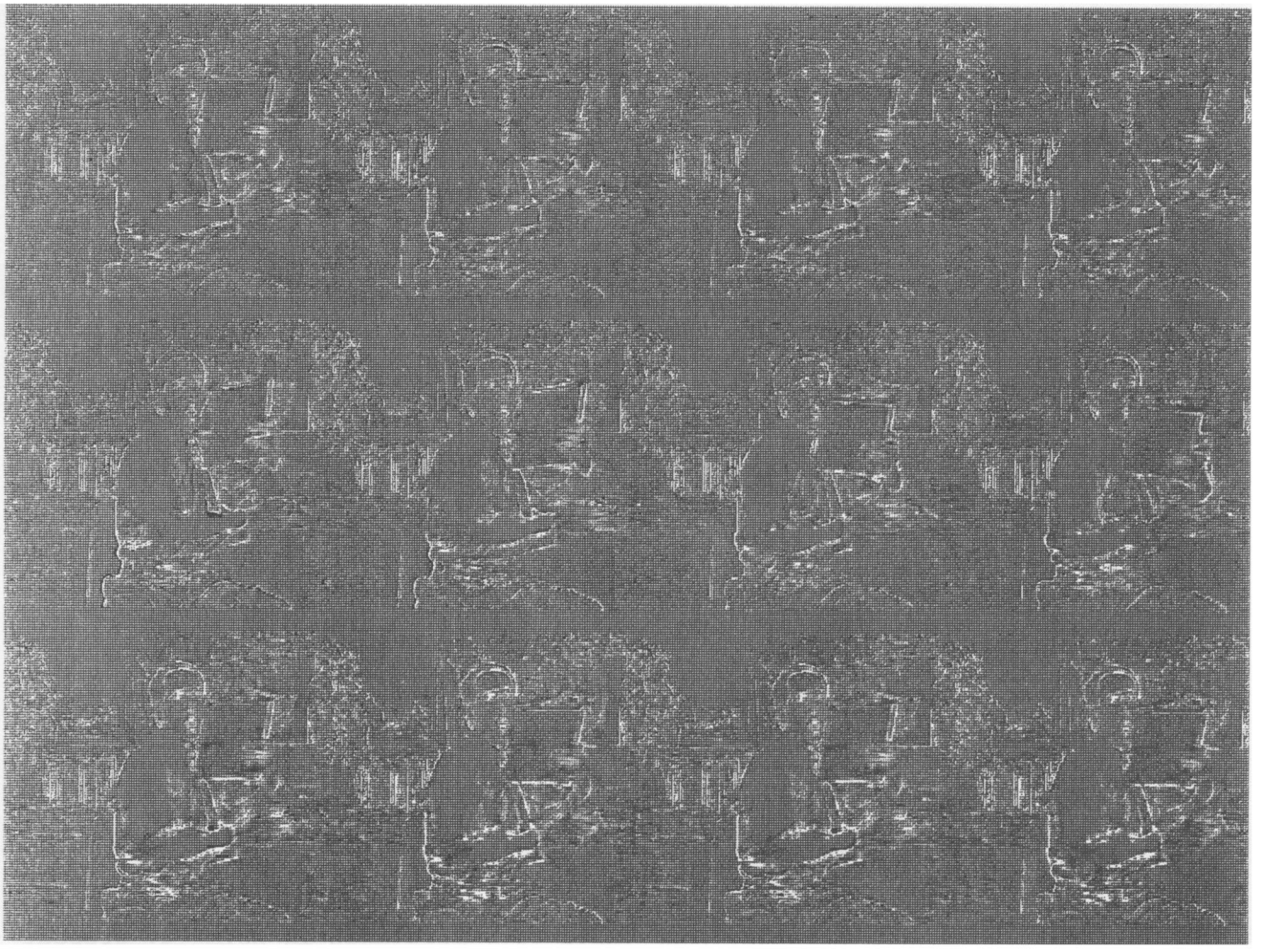

Figure 13. Sobel difference imagery of Figure 7. Top row – rate DS1. Second row – rate 1/2 DS1. Third row – rate 1/4 DS1.

Table 6. Spatial Blurring Features For Figure 13

Scene	M-PSDI	SD-PSDI	RMS-PSDI	NPGT-PSDI
Top Row (DS1)	17.5	35.2	39.4	5443
Second Row (1/2 DS1)	19.0	37.9	42.4	6562
Bottom Row (1/4 DS1)	22.5	47.4	52.5	9544

Table 7. False Edge Features For Figure 13

Scene	M-NSDI	SD-NSDI	RMS-NSDI	NPLT-NSDI
Top Row (DS1)	-9.0	20.7	22.6	1138
Second Row (1/2 DS1)	-9.6	22.5	24.4	1577
Bottom Row (1/4 DS1)	-11.5	25.0	27.5	1989

2.7 Jerkiness Feature Using Position Errors

Jerkiness is a video teleconferencing/telephony artifact in which the original smooth and continuous imagery motion is perceived as a series of distinct snapshots at the output (see Table 1). Jerkiness is normally present when a codec data compression algorithm achieves data compression by elimination of fields or frames. The number of fields and/or frames that are eliminated (not transmitted) is not necessarily guaranteed to be an accurate measure of jerkiness. Sophisticated coding algorithms can update different portions of the image at different frame rates and even interpolate missing frames to achieve smooth motion effects. Jerkiness is present when the position of a moving object within the video scene is not updated rapidly enough. Section 2.7.1 proposes a measure for jerkiness based on injecting a video scene containing a moving object, and then measuring the object's position